A Novel Approach for Mining Trajectory Patterns of Moving Vehicles

Vaishali Mirge
Department of computer
Science & Engineering
M. M. College of Technology,
Raipur (C.G.)

Shubhrata Gupta
Department of Electrical
Engineering
National Institute of
Technology, Raipur (C.G.)

Kesari Verma
Department of Computer
Applications
National Institute of
Technology, Raipur (C.G.)

ABSTRACT

With the advances in location- acquisition technologies such as Global Positioning System (GPS), Global System for Mobile Communications (GSM) etc, increasing amounts of movement data collected from various moving objects such as animals, vehicles, mobile devices, and climate radars have become widely available. Turning a collection of time-geography data into mobility knowledge is a key issue in many research domain. In this article we focus on the analysis of trajectories of moving vehicles and identifying the paths on road networks, which are suffering with heavy traffic. This paper proposes a novel algorithm to get trajectory patterns as a sequence of spatio-temporal regions. These sequences specify the paths heavily loaded with vehicles in certain duration. The problem domain is divided into two parts, First in order to discover spatio-temporal regions, the data space is partitioned into meaningful clusters called spatio-temporal regions. Second, arranged the spatio-temporal regions according to ascending order of time to obtain the trajectory patterns, which shows the paths frequently followed by most of the vehicles in certain duration i.e. path suffered with heavy traffic.

General Terms

Trajectory Pattern Mining, Mobility Data Mining

Keywords

Trajectory Pattern, Traffic control, Mobility Mining, Trajectory Data clustering

1. INTRODUCTION

With the fast development of positioning technology, massive amounts of object movement data have been collected from various moving objects, such as animals, mobile devices, vehicles, and climate radars. As moving object data is widely available, mining and understanding such data has gained a lot of attention recently. The phenomenon of changing the geometrical position of moving objects over time is represented in the form of trajectories. Huge volume of trajectory data provides us an opportunity and challenge to discover automatically novel and useful patterns.

This paper precisely address the problem of extracting trajectory patterns and introduces a novel pattern mining algorithm to get trajectory patterns which is represented by the sequence of spatio-temporal regions , followed by many vehicles in a certain duration i.e. we are identifying the paths heavily dense on road network in certain duration. The proposed algorithm is comprises of two steps: First, cluster the trajectory data into disjoint regions, composite of spatial and temporal values. The concept of density based clustering method called DBSCAN [4] is adopted for this purpose. In the proposed algorithm trajectory points of different trajectories those are satisfying closeness property[1] with respect to location and time are grouped together to form clusters. All these clusters are treated as the spatio-temporal regions. Finally identified the sequence of spatio-temporal regions denoting the path or sub path, which have been followed most by different vehicles in certain duration. This has been achieved by, finding the spatio-temporal regions, for each path, containing number of trajectory points more than a given traffic threshold. Arranging these spatio-temporal regions according to time, will provide us the sequences of spatio-temporal regions denoting the path or sub path with heavy traffic.

The remainder of the paper is organized as follow : Section 2 reviews the related literature. Section 3 comprises of the formal definition of trajectory patterns. Section 4 presents proposed method in detail. Section 5 explains the algorithm 1 and algorithm 2 with an example. In section 6 we present an experimental evaluation of our approach. Finally, Section 7 concludes this paper.

Lemma[1]- Closeness Property

Trajectory point (x_i, y_i, t_i) of T_i and (x_k, y_k, t_k) of T_k are said to be close to each other if

$$\sqrt{(x_i-x_k)^2+(y_i-y_k)^2} \leq \sigma \text{ and } |t_i-t_k| < \varepsilon$$,Where σ and ε are distance and temporal threshold.

2. RELATED WORK

Existing studies on mining spatio-temporal patterns can be generally divided into two categories, according to the types of input data. [1] Discovers patterns from historical datasets of geosciences to detect significant environmental events such as temperature changes. Since we are interested in analyzing the movement patterns of objects, we focus on the mining trajectory data of moving objects. In [3] clustering of spatial data is based on the distance of the points of a cluster to their nearest neighbors. It analyzes the expected distribution of these distances for a cluster. The analysis in this paper is based on the assumption that the points inside of a cluster are uniformly distributed. [5] MoveMine, integrates many data mining functions including moving object pattern mining and trajectory mining based on state-of-the-art methods. It can automatically detect an approximate period in movements, and reveal collective movement patterns like flocks,

followers, and swarms. It can be used to perform trajectory clustering, classification and outlier detection for geometric analysis of trajectories. [7] Introduced the trajectory pattern mining problem, together with several different methods to extract T-patterns from trajectory data. It presents, T-patterns as a basic building block for spatio-temporal data mining, around which more sophisticated analysis tools can be constructed. [8] proposes a direct discriminative pattern mining approach, DDPMine, to tackle the efficiency issue. DDPMine performs a branch-and bound search for directly mining discriminative patterns without generating the complete pattern set. Instead of selecting best patterns in a batch, It introduce a feature-centered mining approach that generates discriminative patterns sequentially on a progressively shrinking FP-tree by incrementally eliminating training instances. The instance limitation effectively reduces the problem size iteratively and expedites the mining process. Wang et. al. [9] proposes a hierarchical statistical information grid based approach for spatial data mining to reduce the cost of scanning spatial data to answer any query. The idea is to capture statistical information associated with spatial cells in such a manner that whole classes of queries and clustering problems can be answered without recourse to the individual objects. Kang et. al. [10] presented an approach for mining spatio-temporal patterns from trajectory data. This paper introduced the problem of representing spatio-temporal properties with redundant location symbols and proposed an efficient method for mining spatio-temporal frequent patterns. It discovers spatio-temporal regions using line simplification and clustering, and extracts spatio-temporal frequent patterns in prefix-projection approach.

3. PROBLEM DEFINITION

The Purpose of work is to extract the paths in road networks which are suffered with heavy traffic. Problem domain is divided in two Parts, First to cluster the trajectory data in order to find dense regions then Second to arrange these clusters according to temporal values to discover trajectory Patterns.

This section introduces some basic concepts and terminology also. Here briefly presents some definitions like trajectory, trajectory patterns and ROI.

Definition 1 (Trajectory)
A trajectory is a trace generated by a moving object in geographical spaces, usually represented by a series of chronologically ordered points, $\{(x_0,y_0,t_0),...,(x_n, y_n, t_n)\}$ where $(x_i, y_i,)$ are positions of moving objects, t_i is a time stamp, for $i = 0,1,...,n$ and $t_0 < t_1 < ... < t_n$.

Definition 2 (Trajectory Pattern)

A Trajectory Pattern (T-pattern) is a couple (S, α) : Where $S = \langle(x_0, y_0),..., (x_k, y_n)\rangle$ is a sequence of n+1 locations, and $\alpha = \langle\alpha_1,..., \alpha_n\rangle$ are the transition times.

A T-pattern T_p occurs in a trajectory if the trajectory contains a subsequence S such that : Each (x_i, y_i) in T_p matches a point (x_i', y_i') in S, and the transition times in T_p are similar to those in S.

Definition 3 (Regions-of-Interest (ROI))

It represents a natural way to partition the space into meaningful areas and correspondingly, to associate spatial points with region labels.

4. PROPOSED METHOD
In order to mine trajectory pattern from moving vehicle trajectories, the problem domain in divided in two parts

1. Clustering the collection of time-geography data into meaningful spatio-temporal regions.

2. Finally obtain the sequence of the spatio-temporal regions to determine the paths heavily loaded with traffic.

4.1 Discovering Spatio-Temporal Regions

In order to discover trajectory patterns , First partition the data space into spatiotemporally meaningful regions. Each region consists of trajectory points those are close to each other with respect to distance & time. Clustering of the time space data into the disjointed spatio-temporal regions is done by using DBSCAN method [4]. In DBSCAN method points are segmented into different clusters according to the density of space. In our approach for clustering the trajectory points, time is also considered as a factor.

Let $D_T = \{T_1, T_2, T_3 ... T_n\}$ be a database of trajectories. Where each T_i is a sequence of triples (x_i, y_i, t_i) for i=1,2...n. For clustering spatio-temporal data, start with the first cluster R_0 with a first point (x_{0i}, y_{0i}, t_{0i}) of any randomly selected Trajectory T_i. Now compare this point with the points (x_j, y_j, t_j) of all other trajectories and points of trajectory T_i, except the point (x_{0i}, y_{0i}, t_{0i}), such that . $\sqrt{(x_j-x_{pi})^2+(y_j-y_{pi})^2} \leq \sigma$ and $|t_j-t_{pi}| < \varepsilon$,Where σ and ε are threshold with respect to distance and time. All such points will be added to the cluster R_0. The same process will be repeated for all other points in the cluster R_0 and if get the points satisfying the above conditions then we add these points also to the cluster R_0. Now cluster R_0 is having all the points which are closer to each with respect to distance and time threshold. Same process will be repeated with all other points not in R_0 and remaining clusters will be formed. The collection of such clusters are treated as a spatio-temporal regions.

Algorithm 1: Discovering Spatio-Temporal regions

Input : Trajectory Database D_T ={T_1, T_2, T_3 ... Tm} where Trajectory T_i = {id, pathid, (x_{i1},y_{i1},t_{i1}),(x_{i1},y_{i2},t_{i2}),…,(x_{im},y_{im},t_{im})}, in which id & pathid are the id of trajectory and its path respectively. Distance threshold σ, Temporal threshold ε, Traffic Threshold α.

Output : A set of spatio-temporal regions.

Procedure :

    Begin

1.   Arrange all trajectory points in the ascending order of time for each trajectory.
2.   Set k=-1, s=0 and R_0 as first cluster.
3.   **For each** trajectory $T_i \in D_T$ where i=1,2,...,m {   **//T_i is a trajectory in D_T**
4.   **For each** point p of T_i {
5.   **If** p is not in any cluster formed.   **// Allot cluster to point p**
5.1   k=k+1 , s=k and add p to R_s .

          Else

5.2   s = cluster number which contains p.
6.   **For each** Trajectory T_j where j=1,2...,m {   **//Cluster Construction**
7.   **If** pathid of T_i and T_j are same then
7.1   **For each** $(x_{qj}, y_{qj}, t_{qj}) \in T_j$ (except p of T_i when j=i) {

          // checking closeness property as Lemma[1]

$$\textbf{If } (\sqrt{(x_{qj} - x_{pi})^{2} + (y_{qj}\text{-}y_{pi})^2} \leq \sigma \text{ and } |t_{qj}\text{-}t_{pi}| < \varepsilon)$$

7.1.1.   Add point (x_{qj}, y_{qj}, t_{qj}) to cluster R_s .

          Else

7.1.2.   Choose next T_j and goto step 6.

          }

        }

      }

    }

8.   **If** number of points in any cluster < α, Reject that cluster.

    End.

4.2 Mining Trajectory Pattern

The collection of spatio-temporal regions has been obtained from the original trajectories using algorithm 1. Now trajectory patterns can be discovered by using these spatio-temporal regions. In order to discover the sequence of spatio-temporal regions followed most, set of spatio-temporal if exists, for each path are arranged according to ascending order of time. These sequences shows the path or sub path busy for the certain duration.

Algorithm 2 : Mining Trajectory Pattern

Input: Set of spatio-temporal regions R={R_1,R_2,..,.R_n} obtained from Algoritm1.

Output: Trajectory Pattern

Procedure :

  Begin

    For each pathid {

        If spatio-temporal regions exists

            Find average of all temporal values inside each spaio-temporal regions..

         Arrange spatio-temporal regions in the ascending order of average time.

         The sequence of spatio-temporal regions represents trajectory pattern for the specified pathid. This trajectory pattern actually shows path or sub path which is having heavy traffic.
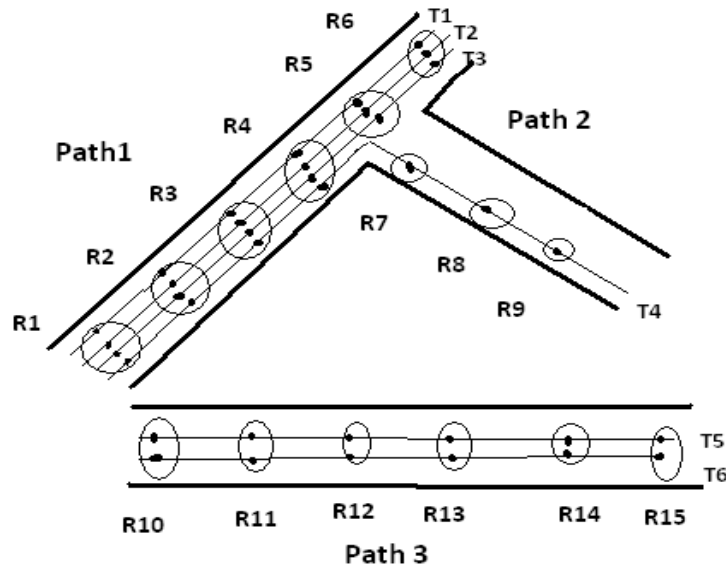
    }

  End.

5. EXAMPLE

Fig 1: Trajectories of Moving Vehicles

In the example given in Fig 1, T_1, T_2, T_3, T_4, T_5, and T_6 are six trajectories. Here R_1, R_2, … R_{15} are clusters formed by combining trajectory points according to algorithm 1. Here trajectories T_1, T_2, T_3, T_4 are in same path, similarly T_5 and T_6 also. After certain duration T_4 has changed the path. When applying clustering algorithm as mentioned in algorithm 1, only the trajectory points on the same path will be considered for clustering. According to algorithm 1 first of all trajectory points of path 1 are taken for clustering. Now by applying algorithm 1, due to closeness property as mentioned in Lemma[1], The clusters R_1, R_2, R_3, R_4, R_5, R_6 can be obtained. In all these six clusters points are close to each other with respect to distance threshold σ and temporal threshold ε. Now the same process will be repeated for Path 2 and Path 3 and get the clusters R_7, R_8, R_9 and R_{10}, R_{11}, R_{12} R_{13}, R_{14}, R_{15} respectively. If set the traffic threshold $\alpha=3$, which shows the traffic density, then the clusters R_7, R_8, R_9 and R_{10}, R_{11}, R_{12} R_{13}, R_{14}, R_{15} are discarded due to they have less than three vehicles. In this way discard all the clusters having trajectory points less than traffic threshold α. Now the clusters R_1, R_2, R_3, R_4, R_5, R_6 are available for algorithm 2. By applying algorithm 2, clusters are arranged in ascending order of time, for path1 and we get the sequence $R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_4 \rightarrow R_5 \rightarrow R_6$ which shows the path or sub path having heavy traffic. Duration of traffic can be obtained by taking the difference between average of temporal values

.

in cluster R_1 and average of temporal values in cluster R_6. If set the traffic threshold $\alpha=4$ the sequence $R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_4$ will show sub-path having heavy traffic.

6. EXPERIMENT AND RESULTS

6.1 Synthetic Data

To evaluate our framework, we implemented our proposal in c++ programming language. Experiments were performed in Pentium IV machine with 40 GB Hard disk and 1 GB RAM. We generated the synthetic data to facilitate performance study. Data has been tested for different values of Distance threshold σ, Temporal threshold ε and traffic threshold α. Different outcomes of the experiment were shown in fig 2.

6.2 Performance Measure

Through the experiments it has been observed that if we increase the traffic threshold we will get fewer numbers of frequent patterns. Increasing the traffic threshold reduces the number of clusters, since clusters having number of trajectory points less than traffic threshold are discarded. Reduction in number of clusters reduces the number of frequent patterns also. Fig 2(a) shows this scenario. In Fig 2(b) it has been represented that if we increase distance threshold, number of clusters formed will be reduced. Increment in distance threshold increases the number of trajectory points inside each clusters and this will reduce the number of clusters formed
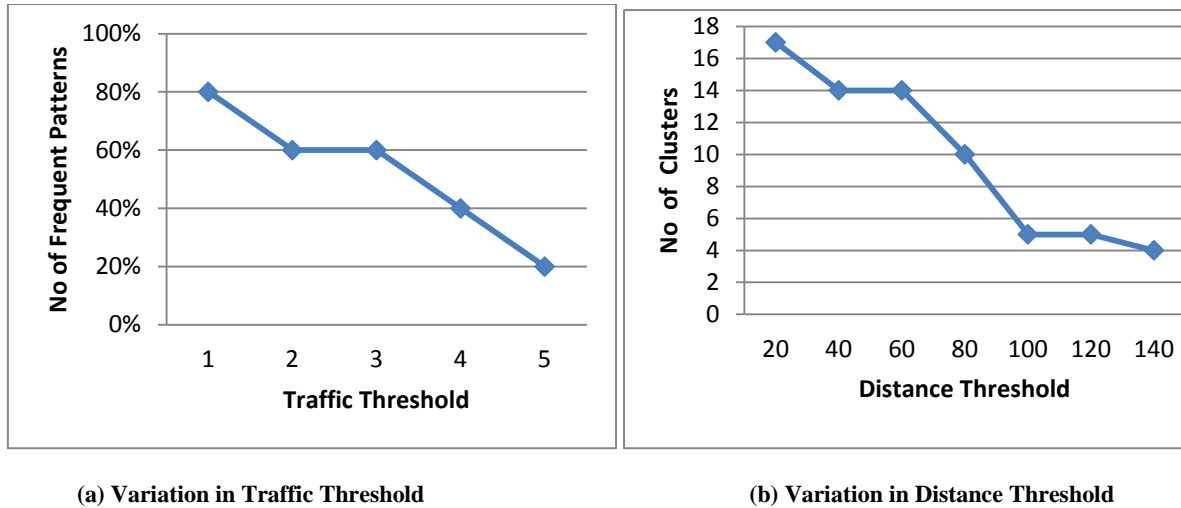
(a) Variation in Traffic Threshold

(b) Variation in Distance Threshold

Fig 2: Effect of Threshold Variation

7. CONCLUSIONS

In this paper, we presented an approach for mining trajectory pattern for the moving vehicles on road networks. It is difficult to deal with spatial locations. In order to deal with this problem, an efficient method for clustering spatio-temporal data into spatio-temporal regions has been proposed. The whole problem of trajectory pattern mining is divided into two parts First generated spatio-temporal regions from trajectory data. For this purpose concept from DBSCAN [4] method is adopted by considering the temporal values also . Only the spatio-temporal regions having trajectory points greater than the traffic threshold are considered for algorithm 2. Secondly for a particular path the spatio-temporal regions, if exists, are arranged according to ascending order of time in order to get the trajectory patterns. Such sequences obtained show the paths or sub-paths having heavy traffic.

8. REFERENCES

[1] Tsoukatos, I. and Gunopulos, D. 2001. Efficient Mining Spatio Temporal Pattern. Proceedings of International Symposium in Spatial and Temporal Databases.425-442.

[2] Cheng, H., Yan X., Han J. Philip S. 2008. Direct Discriminative Pattern Mining for Effective Classification. ICDE '08 Proceedings of the 2008 IEEE 24th International Conference on Data Engineering.

[3] Xiaowei, X., Martin, E., Kriegel, H., and Sander, J. 1998. A Distribution-Based clustering Algorithm for Mining in Large Spatial Databases. Proceedings of 14th International Conference on Data Engineering (ICDE).169-178.

[4] Ester M., Kriegel H., Sandar J., Xu X. and Kut, A. 1998. A Density-Based Clustering in Spatial Databases : The Algorithm GDBSCAN and Its Applications. Data Mining & Knowledge Discovery. Volume 2. Issue 2.169-194.

[5] Li,Z., Ji,M., Lee, J-G.,Tang, L., Han, J., and Kays, R..2010. MoveMine: Mining Moving Object Databases. International Conference on Management of Data.

[6] Agrawal R. and Srikant, R. 1995. Mining Sequential Patterns. Proc. 1995 International Conference on Data Engineering (ICDE). 3-14.

[7] Giannotti F., M. Nanni, Pedreschi, D. and Pinelli, F. 2007.Trajectory Pattern Mining, KDD˙'07, August 12-15, San Jose, California, USA. 330-339

[8] H. Cheng, X. Yan, J. Han, and S. Philip, "Direct Discriminative Pattern Mining for Effective Classification", ICDE '08 Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, 2008.

[9] W. Wang, J. Yang, and R. Muntz, "STING: A Statistical Information Grid Approach to Spatial Data Mining", In Proc. 23rd Int'l Conf. on Very Large DataBases, Athens, Greece, pp. 186-195, Aug 1997.

[10] J. Kang and H. Yong "Mining Trajectory Patterns by Incorporating Temporal Properties", In Proceedings of the 1st International Conference on Emerging Databases (EDB2009), 2009.